

24745

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
12 July 2001 (12.07.2001)

PCT

(10) International Publication Number  
WO 01/49864 A1

(51) International Patent Classification<sup>7</sup>: C12N 15/57 (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(21) International Application Number: PCT/US01/00548 (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(22) International Filing Date: 5 January 2001 (05.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 60/174,686 6 January 2000 (06.01.2000) US

(71) Applicant: LEXICON GENETICS INCORPORATED [US/US]; 4000 Research Forest Drive, The Woodlands, TX 77381 (US).

(72) Inventors: WALKE, D., Wade; 7507 Danehill Drive, Spring, TX 77389 (US). WILGANOWSKI, Nathaniel, L.; 9820 Memorial, Apt. 77, Houston, TX 77024 (US). DONOHO, Gregory; 95 Autumn Branch Drive, The Woodlands, TX 77382 (US). TURNER, C., Alexander, Jr.; 67 Winter Wheat Place, The Woodlands, TX 77381 (US).

(74) Agents: ISHIMOTO, Lance, K. et al.; Lexicon Genetics Incorporated, 4000 Research Forest Drive, The Woodlands, TX 77381 (US).



WO 01/49864 A1

## Published:

- With international search report.
- Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL HUMAN PROTEASES AND POLYNUCLEOTIDES ENCODING THE SAME

(57) Abstract: Novel human polynucleotide and polypeptide sequences are disclosed that can be used in therapeutic, diagnostic, and pharmacogenomic applications.

NOVEL HUMAN PROTEASES AND POLYNUCLEOTIDES  
ENCODING THE SAME

The present application claims the benefit of U.S. 5 Provisional Application Number 60/174,686 which was filed on January 6, 2000 and is herein incorporated by reference in its entirety.

## 1. INTRODUCTION

The present invention relates to the discovery, 10 identification, and characterization of novel human polynucleotides encoding proteins sharing sequence similarity with mammalian proteases. The invention encompasses the described polynucleotides, host cell expression systems, the encoded proteins, fusion proteins, polypeptides and peptides, 15 antibodies to the encoded proteins and peptides, and genetically engineered animals that either lack or over express the disclosed sequences, antagonists and agonists of the proteins, and other compounds that modulate the expression or activity of the proteins encoded by the disclosed 20 polynucleotides that can be used for diagnosis, drug screening, clinical trial monitoring and the treatment of physiological disorders.

## 2. BACKGROUND OF THE INVENTION

25 Proteases cleave protein substrates as part of degradation, maturation, and secretory pathways within the body. Proteases have been associated with, *inter alia*, regulating development, infertility, modulating cellular processes, fertility, and infectious disease.

30

## 3. SUMMARY OF THE INVENTION

The present invention relates to the discovery, 35 identification, and characterization of nucleotides that encode novel human proteins, and the corresponding amino acid sequences of these proteins. The novel human proteins (NHPs) described for the first time herein share structural similarity with animal proteases, and particularly trypsin-

like serine proteases such as enteropeptidase (enterokinase), plasminogen, and acrosin.

The novel human nucleic acid (cDNA) sequences described herein, encode proteins/open reading frames (ORFs) of 217, 5 348, and 288 amino acids in length (see SEQ ID NOS: 2, 4, and 6 respectively).

The invention also encompasses agonists and antagonists of the described NHPs, including small molecules, large molecules, mutant NHPs, or portions thereof that compete with 10 native NHPs, NHP peptides, and NHP antibodies, as well as nucleotide sequences that can be used to inhibit the expression of the described NHPs (e.g., antisense and ribozyme molecules, and gene or regulatory sequence replacement constructs) or to enhance the expression of the described NHPs 15 (e.g., expression constructs that place the described sequence under the control of a strong promoter system), and transgenic animals that express a NHP transgene, or "knock-outs" (which can be conditional) that do not express a functional NHP.

Further, the present invention also relates to processes 20 for identifying compounds that modulate, i.e., act as agonists or antagonists, of NHP expression and/or NHP activity that utilize purified preparations of the described NHP and/or NHP product, or cells expressing the same. Such compounds can be used as therapeutic agents for the treatment of any of a wide 25 variety of symptoms associated with biological disorders or imbalances.

#### 4. DESCRIPTION OF THE SEQUENCE LISTING AND FIGURES

The Sequence Listing provides the sequences of the NHP 30 ORFs encoding the described NHP amino acid sequences. SEQ ID NO: 7 describes a NHP ORF with flanking sequences.

#### 5. DETAILED DESCRIPTION OF THE INVENTION

The NHPs, described for the first time herein, are novel 35 proteins that are expressed in, *inter alia*, human cell lines, and human testis cells.

The described sequences were compiled from gene trapped cDNAs and clones isolated from a human testis cDNA library (Edge Biosystems, Gaithersburg, MD). The present invention encompasses the nucleotides presented in the Sequence Listing, 5 host cells expressing such nucleotides, the expression products of such nucleotides, and: (a) nucleotides that encode mammalian homologs of the described sequences, including the specifically described NHPs, and the NHP products; (b) nucleotides that encode one or more portions of a NHP that 10 correspond to functional domains of the NHP, and the polypeptide products specified by such nucleotide sequences, including but not limited to the novel regions of any active domain(s); (c) isolated nucleotides that encode mutant versions, engineered or naturally occurring, of a described 15 NHP in which all or a part of at least one domain is deleted or altered, and the polypeptide products specified by such nucleotide sequences, including but not limited to soluble proteins and peptides in which all or a portion of the signal sequence is deleted; (d) nucleotides that encode chimeric 20 fusion proteins containing all or a portion of a coding region of a NHP, or one of its domains (e.g., a receptor or ligand binding domain, accessory protein/self-association domain, etc.) fused to another peptide or polypeptide; or (e) therapeutic or diagnostic derivatives of the described 25 polynucleotides such as oligonucleotides, antisense polynucleotides, ribozymes, dsRNA, or gene therapy constructs comprising a sequence first disclosed in the Sequence Listing.

As discussed above, the present invention includes:

(a) the human DNA sequences presented in the Sequence Listing 30 (and vectors comprising the same) and additionally contemplates any nucleotide sequence encoding a contiguous NHP open reading frame (ORF), or a contiguous exon splice junction first described in the Sequence Listing, that hybridizes to a complement of a DNA sequence presented in the Sequence Listing 35 under highly stringent conditions, e.g., hybridization to filter-bound DNA in 0.5 M NaHPO<sub>4</sub>, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1xSSC/0.1% SDS at

68°C (Ausubel F.M. et al., eds., 1989, *Current Protocols in Molecular Biology*, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York, at p. 2.10.3) and encodes a functionally equivalent gene product. Additionally 5 contemplated are any nucleotide sequences that hybridize to the complement of the DNA sequence that encode and express an amino acid sequence presented in the Sequence Listing under moderately stringent conditions, e.g., washing in 0.2xSSC/0.1% SDS at 42°C (Ausubel et al., 1989, *supra*), yet still encode a 10 functionally equivalent NHP product. Functional equivalents of a NHP include naturally occurring NHPs present in other species and mutant NHPs whether naturally occurring or engineered (by site directed mutagenesis, gene shuffling, directed evolution as described in, for example, U.S. Patent 15 No. 5,837,458). The invention also includes degenerate nucleic acid variants of the disclosed NHP polynucleotide sequences.

Additionally contemplated are polynucleotides encoding a NHP ORF, or its functional equivalent, encoded by a 20 polynucleotide sequence that is about 99, 95, 90, or about 85 percent similar or identical to corresponding regions of the nucleotide sequences of the Sequence Listing (as measured by BLAST sequence comparison analysis using, for example, the GCG sequence analysis package using standard default settings).

25 The invention also includes nucleic acid molecules, preferably DNA molecules, that hybridize to, and are therefore the complements of, the described NHP nucleotide sequences. Such hybridization conditions may be highly stringent or less highly stringent, as described above. In instances where the 30 nucleic acid molecules are deoxyoligonucleotides ("DNA oligos"), such molecules are generally about 16 to about 100 bases long, or about 20 to about 80, or about 34 to about 45 bases long, or any variation or combination of sizes represented therein that incorporate a contiguous region of 35 sequence first disclosed in the Sequence Listing. Such oligonucleotides can be used in conjunction with the

polymerase chain reaction (PCR) to screen libraries, isolate clones, and prepare cloning and sequencing templates, etc.

Alternatively, such NHP oligonucleotides can be used as hybridization probes for screening libraries, and assessing 5 gene expression patterns (particularly using a micro array or high-throughput "chip" format). Additionally, a series of the described NHP oligonucleotide sequences, or the complements thereof, can be used to represent all or a portion of the described NHP sequences. An oligonucleotide or polynucleotide 10 sequence first disclosed in at least a portion of one or more of the sequences of SEQ ID NOS: 1-7 can be used as a hybridization probe in conjunction with a solid support matrix/substrate (resins, beads, membranes, plastics, polymers, metal or metallized substrates, crystalline or 15 polycrystalline substrates, etc.). Of particular note are spatially addressable arrays (*i.e.*, gene chips, microtiter plates, etc.) of oligonucleotides and polynucleotides, or corresponding oligopeptides and polypeptides, wherein at least one of the biopolymers present on the spatially addressable 20 array comprises an oligonucleotide or polynucleotide sequence first disclosed in at least one of the sequences of SEQ ID NOS: 1-7, or an amino acid sequence encoded thereby. Methods for attaching biopolymers to, or synthesizing biopolymers on, solid support matrices, and conducting binding studies thereon 25 are disclosed in, *inter alia*, U.S. Patent Nos. 5,700,637, 5,556,752, 5,744,305, 4,631,211, 5,445,934, 5,252,743, 4,713,326, 5,424,186, and 4,689,405 the disclosures of which are herein incorporated by reference in their entirety.

Addressable arrays comprising sequences first disclosed 30 in SEQ ID NOS:1-7 can be used to identify and characterize the temporal and tissue specific expression of a sequence. These addressable arrays incorporate oligonucleotide sequences of sufficient length to confer the required specificity, yet be within the limitations of the production technology. The 35 length of these probes is within a range of between about 8 to about 2000 nucleotides. Preferably the probes consist of 60

nucleotides and more preferably 25 nucleotides from the sequences first disclosed in SEQ ID NOS:1-7.

For example, a series of the described oligonucleotide sequences, or the complements thereof, can be used in chip 5 format to represent all or a portion of the described sequences. The oligonucleotides, typically between about 16 to about 40 (or any whole number within the stated range) nucleotides in length can partially overlap each other and/or the sequence may be represented using oligonucleotides that do 10 not overlap. Accordingly, the described polynucleotide sequences shall typically comprise at least about two or three distinct oligonucleotide sequences of at least about 8 nucleotides in length that are each first disclosed in the described Sequence Listing. Such oligonucleotide sequences 15 can begin at any nucleotide present within a sequence in the Sequence Listing and proceed in either a sense (5'-to-3') orientation vis-a-vis the described sequence or in an antisense orientation.

Microarray-based analysis allows the discovery of broad 20 patterns of genetic activity, providing new understanding of gene functions and generating novel and unexpected insight into transcriptional processes and biological mechanisms. The use of addressable arrays comprising sequences first disclosed in SEQ ID NOS:1-7 provides detailed information about 25 transcriptional changes involved in a specific pathway, potentially leading to the identification of novel components or gene functions that manifest themselves as novel phenotypes.

Probes consisting of sequences first disclosed in SEQ ID 30 NOS:1-7 can also be used in the identification, selection and validation of novel molecular targets for drug discovery. The use of these unique sequences permits the direct confirmation of drug targets and recognition of drug dependent changes in gene expression that are modulated through pathways distinct 35 from the drugs intended target. These unique sequences therefore also have utility in defining and monitoring both drug action and toxicity.

As an example of utility, the sequences first disclosed in SEQ ID NOS:1-7 can be utilized in microarrays or other assay formats, to screen collections of genetic material from patients who have a particular medical condition. These investigations can also be carried out using the sequences previously collected genetic databases and the disclosed sequences using computer software known to those in the art. Thus the sequences first disclosed in SEQ ID NOS:1-7 in silico and by comparing sequences to identify mutations associated with a particular disease and also as a diagnostic or prognostic assay.

Although the presently described sequences have been specifically described using nucleotide sequence, it should be appreciated that each of the sequences can uniquely be described using any of a wide variety of additional structural attributes, or combinations thereof. For example, a given sequence can be described by the net composition of the nucleotides present within a given region of the sequence in conjunction with the presence of one or more specific oligonucleotide sequence(s) first disclosed in the SEQ ID NOS:1-7. Alternatively, a restriction map specifying the relative positions of restriction endonuclease digestion sites, or various palindromic or other specific oligonucleotide sequences can be used to structurally describe a given sequence. Such restriction maps, which are typically generated by widely available computer programs (e.g., the SEQUENCER 3.0, Gene Codes Corp., Ann Arbor, MI, etc.), can optionally be used in conjunction with one or more discrete nucleotide sequence(s) present in the sequence that can be described by the relative position of the sequence relative to one or more additional sequence(s) or one or more restriction sites present in the disclosed sequence.

For oligonucleotide probes, highly stringent conditions may refer, e.g., to washing in 6xSSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligos), 48°C (for 17-base oligos), 55°C (for 20-base oligos), and 60°C (for 23-base

oligos). These nucleic acid molecules may encode or act as NHP gene antisense molecules, useful, for example, in NHP gene regulation (for and/or as antisense primers in amplification reactions of NHP nucleic acid sequences). With respect to NHP 5 gene regulation, such techniques can be used to regulate biological functions. Further, such sequences may be used as part of ribozyme and/or triple helix sequences that are also useful for NHP gene regulation.

Inhibitory antisense or double stranded oligonucleotides 10 can additionally comprise at least one modified base moiety which is selected from the group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-15 2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 20 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiacytosine, 25 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)<sup>w</sup>, and 2,6-diaminopurine.

The antisense oligonucleotide can also comprise at least 30 one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the antisense oligonucleotide will comprise at least one modified phosphate backbone 35 selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate,

a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

In yet another embodiment, the antisense oligonucleotide is an  $\alpha$ -anomeric oligonucleotide. An  $\alpha$ -anomeric 5 oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gautier et al., 1987, Nucl. Acids Res. 15:6625-6641). The oligonucleotide is a 2'-0-methylribonucleotide (Inoue et al., 1987, Nucl. Acids Res. 10 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330). Alternatively, double stranded RNA can be used to disrupt the expression and function of a targeted NHP.

Oligonucleotides of the invention can be synthesized by 15 standard methods known in the art, e.g. by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate oligonucleotides can be synthesized by the method of Stein et al. (1988, Nucl. Acids Res. 16:3209), and 20 methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:7448-7451), etc.

Low stringency conditions are well known to those of skill in the art, and will vary predictably depending on the 25 specific organisms from which the library and the labeled sequences are derived. For guidance regarding such conditions see, for example, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual (and periodic updates thereof), Cold Springs Harbor Press, N.Y.; and Ausubel et al., 1989, Current 30 Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y.

Alternatively, suitably labeled NHP nucleotide probes can be used to screen a human genomic library using appropriately stringent conditions or by PCR. The identification and 35 characterization of human genomic clones is helpful for identifying polymorphisms (including, but not limited to, nucleotide repeats, microsatellite alleles, single nucleotide

polymorphisms, or coding single nucleotide polymorphisms), determining the genomic structure of a given locus/allele, and designing diagnostic tests. For example, sequences derived from regions adjacent to the intron/exon boundaries of the 5 human gene can be used to design primers for use in amplification assays to detect mutations within the exons, introns, splice sites (e.g., splice acceptor and/or donor sites), etc., that can be used in diagnostics and pharmacogenomics.

10 Further, a NHP homolog can be isolated from nucleic acid from an organism of interest by performing PCR using two degenerate or "wobble" oligonucleotide primer pools designed on the basis of amino acid sequences within the NHP products disclosed herein. The template for the reaction may be total 15 RNA, mRNA, and/or cDNA obtained by reverse transcription of mRNA prepared from human or non-human cell lines or tissue known or suspected to express an allele of a NHP gene.

20 The PCR product can be subcloned and sequenced to ensure that the amplified sequences represent the sequence of the desired NHP gene. The PCR fragment can then be used to isolate a full length cDNA clone by a variety of methods. For example, the amplified fragment can be labeled and used to screen a cDNA library, such as a bacteriophage cDNA library. Alternatively, the labeled fragment can be used to isolate 25 genomic clones via the screening of a genomic library.

PCR technology can also be used to isolate full length cDNA sequences. For example, RNA can be isolated, following standard procedures, from an appropriate cellular or tissue source (i.e., one known, or suspected, to express a NHP 30 sequence). A reverse transcription (RT) reaction can be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis. The resulting RNA/DNA hybrid may then be "tailed" using a standard terminal transferase 35 reaction, the hybrid may be digested with RNase H, and second strand synthesis may then be primed with a complementary primer. Thus, cDNA sequences upstream of the amplified

fragment can be isolated. For a review of cloning strategies that can be used, see e.g., Sambrook *et al.*, 1989, *supra*.

A cDNA encoding a mutant NHP gene can be isolated, for example, by using PCR. In this case, the first cDNA strand 5 may be synthesized by hybridizing an oligo-dT oligonucleotide to mRNA isolated from tissue known or suspected to be expressed in an individual putatively carrying a mutant NHP allele, and by extending the new strand with reverse transcriptase. The second strand of the cDNA is then 10 synthesized using an oligonucleotide that hybridizes specifically to the 5' end of the normal gene. Using these two primers, the product is then amplified via PCR, optionally cloned into a suitable vector, and subjected to DNA sequence analysis through methods well known to those of skill in the 15 art. By comparing the DNA sequence of the mutant NHP allele to that of a corresponding normal NHP allele, the mutation(s) responsible for the loss or alteration of function of the mutant NHP gene product can be ascertained.

Alternatively, a genomic library can be constructed using 20 DNA obtained from an individual suspected of or known to carry a mutant NHP allele (e.g., a person manifesting a NHP-associated phenotype such as, for example, obesity, high blood pressure, connective tissue disorders, infertility, etc.), or a cDNA library can be constructed using RNA from a tissue 25 known, or suspected, to express a mutant NHP allele. A normal NHP gene, or any suitable fragment thereof, can then be labeled and used as a probe to identify the corresponding mutant NHP allele in such libraries. Clones containing mutant 30 NHP gene sequences can then be purified and subjected to sequence analysis according to methods well known to those skilled in the art.

Additionally, an expression library can be constructed utilizing cDNA synthesized from, for example, RNA isolated from a tissue known, or suspected, to express a mutant NHP 35 allele in an individual suspected of or known to carry such a mutant allele. In this manner, gene products made by the putatively mutant tissue can be expressed and screened using

standard antibody screening techniques in conjunction with antibodies raised against normal NHP product, as described below. (For screening techniques, see, for example, Harlow, E. and Lane, eds., 1988, "Antibodies: A Laboratory Manual", 5 Cold Spring Harbor Press, Cold Spring Harbor.) Additionally, screening can be accomplished by screening with labeled NHP fusion proteins, such as, for example, alkaline phosphatase-NHP or NHP-alkaline phosphatase fusion proteins. In cases where a NHP mutation results in an expressed gene 10 product with altered function (e.g., as a result of a missense or a frameshift mutation), polyclonal antibodies to NHP are likely to cross-react with a corresponding mutant NHP gene product. Library clones detected via their reaction with such 15 labeled antibodies can be purified and subjected to sequence analysis according to methods well known in the art.

The invention also encompasses (a) DNA vectors that contain any of the foregoing NHP coding sequences and/or their complements (i.e., antisense); (b) DNA expression vectors that contain any of the foregoing NHP coding sequences operatively 20 associated with a regulatory element that directs the expression of the coding sequences (for example, baculo virus as described in U.S. Patent No. 5,869,336 herein incorporated by reference); (c) genetically engineered host cells that contain any of the foregoing NHP coding sequences operatively 25 associated with a regulatory element that directs the expression of the coding sequences in the host cell; and (d) genetically engineered host cells that express an endogenous NHP sequence under the control of an exogenously introduced regulatory element (i.e., gene activation). As used herein, 30 regulatory elements include, but are not limited to, inducible and non-inducible promoters, enhancers, operators and other elements known to those skilled in the art that drive and regulate expression. Such regulatory elements include but are not limited to the human cytomegalovirus (hCMV) immediate 35 early gene, regulatable, viral elements (particularly retroviral LTR promoters), the early or late promoters of SV40 adenovirus, the lac system, the trp system, the TAC system,

the TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase (PGK), the promoters of acid phosphatase, and the promoters of the yeast  $\alpha$ -mating factors.

The present invention also encompasses antibodies and anti-idiotypic antibodies (including Fab fragments), antagonists and agonists of a NHP, as well as compounds or nucleotide constructs that inhibit expression of a NHP sequence (transcription factor inhibitors, antisense and ribozyme molecules, or gene or regulatory sequence replacement constructs), or promote the expression of a NHP (e.g., expression constructs in which NHP coding sequences are operatively associated with expression control elements such as promoters, promoter/enhancers, etc.).

The NHPs or NHP peptides, NHP fusion proteins, NHP nucleotide sequences, antibodies, antagonists and agonists can be useful for the detection of mutant NHPs or inappropriately expressed NHPs for the diagnosis of disease. The NHP proteins or peptides, NHP fusion proteins, NHP nucleotide sequences, host cell expression systems, antibodies, antagonists, agonists and genetically engineered cells and animals can be used for screening for drugs (or high throughput screening of combinatorial libraries) effective in the treatment of the symptomatic or phenotypic manifestations of perturbing the normal function of a NHP in the body. The use of engineered host cells and/or animals may offer an advantage in that such systems allow not only for the identification of compounds that bind to the endogenous receptor for a NHP, but can also identify compounds that trigger NHP-mediated activities or pathways.

Finally, the NHP products can be used as therapeutics. For example, soluble derivatives such as NHP peptides/domains corresponding to NHP, NHP fusion protein products (especially NHP-Ig fusion proteins, i.e., fusions of a NHP, or a domain of a NHP, to an IgFc), NHP antibodies and anti-idiotypic antibodies (including Fab fragments), antagonists or agonists

(including compounds that modulate or act on downstream targets in a NHP-mediated pathway) can be used to directly treat diseases or disorders. For instance, the administration of an effective amount of soluble NHP, or a NHP-IgFc fusion protein or an anti-idiotypic antibody (or its Fab) that mimics the NHP could activate or effectively antagonize the endogenous NHP receptor. Nucleotide constructs encoding such NHP products can be used to genetically engineer host cells to express such products *in vivo*; these genetically engineered cells function as "bioreactors" in the body delivering a continuous supply of a NHP, a NHP peptide, or a NHP fusion protein to the body. Nucleotide constructs encoding functional NHP, mutant NHPs, as well as antisense and ribozyme molecules can also be used in "gene therapy" approaches for the modulation of NHP expression. Thus, the invention also encompasses pharmaceutical formulations and methods for treating biological disorders.

Various aspects of the invention are described in greater detail in the subsections below.

20

#### 5.1 THE NHP SEQUENCES

The cDNA sequences (SEQ ID NO: 1, 3, and 5) and the corresponding deduced amino acid sequences of the described NHP are presented in the Sequence Listing. SEQ ID NO:7 describes a NHP ORF as well as flanking regions. The NHP nucleotides were obtained from human cDNA libraries using probes and/or primers generated from human gene trapped sequence tags. Expression analysis has provided evidence that the described NHPs can be expressed in human tissue as well as gene trapped human cells. In addition, the described NHP sequences can contain a variety of polymorphisms such as at nucleotide 28 of SEQ ID NO:3 and nucleotide 55 of SEQ ID NO:3 which both can be a C or a T and can give rise to silent mutation at corresponding amino acid position 10 of SEQ ID NO:4 or a tyr or his at amino acid position 19 of SEQ ID NO:4. The described NHP sequences can also contain G-A polymorphisms at nucleotide 379 of SEQ ID NO:3 and nucleotide position 199

of SEQ ID NO:5 which can give rise to a corresponding ala or thr at amino acid position 127 of SEQ ID NO:4, or residue 67 of SEQ ID NO:6. The described NHPs share similarity with trypsin-like proteases, plasminogens, and acrosins.

5

### 5.2 NHPs AND NHP POLYPEPTIDES

NHPs, polypeptides, peptide fragments, mutated, truncated, or deleted forms of the NHPs, and/or NHP fusion proteins can be prepared for a variety of uses. These uses 10 include, but are not limited to, the generation of antibodies, as reagents in diagnostic assays, for the identification of other cellular gene products related to a NHP, as reagents in assays for screening for compounds that can be as pharmaceutical reagents useful in the therapeutic treatment of 15 mental, biological, or medical disorders and disease.

The Sequence Listing discloses the amino acid sequence encoded by the described NHP polynucleotides. The NHPs display initiator methionines in DNA sequence contexts consistent with a translation initiation site, and display a 20 consensus signal sequence.

The NHP amino acid sequences of the invention include the amino acid sequences presented in the Sequence Listing as well as analogues and derivatives thereof, as well as any oligopeptide sequence of at least about 10-40, generally about 25 12-35, or about 16-30 amino acids in length first disclosed in the Sequence Listing. Further, corresponding NHP homologues from other species are encompassed by the invention. In fact, any NHP encoded by the NHP nucleotide sequences described above are within the scope of the invention, as are any novel 30 polynucleotide sequences encoding all or any novel portion of an amino acid sequence presented in the Sequence Listing. The degenerate nature of the genetic code is well known, and, accordingly, each amino acid presented in the Sequence Listing, is generically representative of the well known 35 nucleic acid "triplet" codon, or in many cases codons, that can encode the amino acid. As such, as contemplated herein, the amino acid sequences presented in the Sequence Listing,

when taken together with the genetic code (see, for example, Table 4-1 at page 109 of "Molecular Cell Biology", 1986, J. Darnell et al. eds., Scientific American Books, New York, NY, herein incorporated by reference) are generically 5 combinations of nucleic acid sequences that can encode such amino acid sequences.

The invention also encompasses proteins that are functionally equivalent to the NHPs encoded by the presently described nucleotide sequences as judged by any of a number of criteria, including, but not limited to, the ability to bind and cleave a substrate of a NHP, or the ability to effect an identical or complementary downstream pathway, or a change in 10 cellular metabolism (e.g., proteolytic activity, ion flux, tyrosine phosphorylation, etc.). Such functionally equivalent NHP proteins include, but are not limited to, additions or substitutions of amino acid residues within the amino acid sequence encoded by the NHP nucleotide sequences described above, but which result in a silent change, thus producing a 15 functionally equivalent gene product. Amino acid and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include 20 glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. 25 A variety of host-expression vector systems can be used to express the NHP nucleotide sequences of the invention. Where, as in the present instance, the NHP products or NHP polypeptides are thought to be soluble or secreted molecules, the peptide or polypeptide can be recovered from the culture media. Such expression systems also encompass engineered host 30 cells that express a NHP, or a functional equivalent, *in situ*.

Purification or enrichment of NHP from such expression systems can be accomplished using appropriate detergents and lipid micelles and methods well known to those skilled in the art. However, such engineered host cells themselves may be used in 5 situations where it is important not only to retain the structural and functional characteristics of the NHP, but to assess biological activity, *e.g.*, in drug screening assays.

The expression systems that may be used for purposes of the invention include but are not limited to microorganisms 10 such as bacteria (*e.g.*, *E. coli*, *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing NHP nucleotide sequences; yeast (*e.g.*, *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing NHP encoding nucleotide 15 sequences; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing NHP sequences; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant 20 plasmid expression vectors (*e.g.*, Ti plasmid) containing NHP nucleotide sequences; or mammalian cell systems (*e.g.*, COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from 25 mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5K promoter).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the NHP product being expressed. For example, when a large 30 quantity of such a protein is to be produced for the generation of pharmaceutical compositions of or containing NHP, or for raising antibodies to a NHP, vectors that direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, 35 but are not limited, to the *E. coli* expression vector pUR278 (Ruther *et al.*, 1983, EMBO J. 2:1791), in which a NHP coding sequence may be ligated individually into the vector in frame

with the *lacZ* coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, 1985, Nucleic Acids Res. 13:3101-3109; Van Heeke & Schuster, 1989, J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors (Pharmacia or 5 American Type Culture Collection) can also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in 10 the presence of free glutathione. The PGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target sequence product can be released from the GST moiety.

In an insect system, *Autographa californica* nuclear 15 polyhidrosis virus (AcNPV) is used as a vector to express foreign sequences. The virus grows in *Spodoptera frugiperda* cells. A NHP coding sequence can be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for 20 example the polyhedrin promoter). Successful insertion of NHP coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to 25 infect *Spodoptera frugiperda* cells in which the inserted sequence is expressed (e.g., see Smith et al., 1983, J. Virol. 46: 584; Smith, U.S. Patent No. 4,215,051).

In mammalian host cells, a number of viral-based 30 expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the NHP nucleotide sequence of interest may be ligated to an adenovirus transcription/translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric sequence may then be inserted in the adenovirus genome by *in* 35 *vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (e.g., region E1 or E3) will result in a recombinant virus that is viable and capable of

expressing a NHP product in infected hosts (e.g., See Logan & Shenk, 1984, Proc. Natl. Acad. Sci. USA 81:3655-3659). Specific initiation signals may also be required for efficient translation of inserted NHP nucleotide sequences. These 5 signals include the ATG initiation codon and adjacent sequences. In cases where an entire NHP gene or cDNA, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in 10 cases where only a portion of a NHP coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure 15 translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression can be enhanced by the inclusion of appropriate transcription enhancer elements, transcription 20 terminators, etc. (See Bittner *et al.*, 1987, Methods in Enzymol. 153:516-544).

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific 25 fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins and 30 gene products. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and 35 phosphorylation of the gene product may be used. Such mammalian host cells include, but are not limited to, CHO,

VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, and in particular, human cell lines.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express the NHP sequences described above can be engineered. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (e.g., promoter, enhancer sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express the NHP product. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the NHP product.

A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler, et al., 1977, Cell 11:223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska & Szybalski, 1962, Proc. Natl. Acad. Sci. USA 48:2026), and adenine phosphoribosyltransferase (Lowy, et al., 1980, Cell 22:817) genes can be employed in tk<sup>-</sup>, hprt<sup>-</sup> or aprt<sup>-</sup> cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for the following genes: dhfr, which confers resistance to methotrexate (Wigler, et al., 1980, Proc. Natl. Acad. Sci. USA 77:3567; O'Hare, et al., 1981, Proc. Natl. Acad. Sci. USA 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan & Berg, 1981, Proc. Natl. Acad. Sci. USA 78:2072); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, et al., 1981, J. Mol.

Biol. 150:1); and hygro, which confers resistance to hygromycin (Santerre, et al., 1984, Gene 30:147).

Alternatively, any fusion protein can be readily purified by utilizing an antibody specific for the fusion protein being expressed. For example, a system described by Janknecht et al. allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (Janknecht, et al., 1991, Proc. Natl. Acad. Sci. USA 88:8972-8976). In this system, the sequence of interest is subcloned into a vaccinia recombination plasmid such that the sequence's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto  $\text{Ni}^{2+}$ ·nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

Also encompassed by the present invention are fusion proteins that direct the NHP to a target organ and/or facilitate transport across the membrane into the cytosol. Conjugation of NHPs to antibody molecules or their Fab fragments could be used to target cells bearing a particular epitope. Attaching the appropriate signal sequence to the NHP would also transport the NHP to the desired location within the cell. Alternatively targeting of NHP or its nucleic acid sequence might be achieved using liposome or lipid complex based delivery systems. Such technologies are described in Liposomes: A Practical Approach, New RRC ed., Oxford University Press, New York and in U.S. Patents Nos. 4,594,595, 5,459,127, 5,948,767 and 6,110,490 and their respective disclosures which are herein incorporated by reference in their entirety. Additionally embodied are novel protein constructs engineered in such a way that they facilitate transport of the NHP to the target site or desired organ, where they cross the cell membrane and/or the nucleus where the NHP can exert its functional activity. This goal may be achieved by coupling of the NHP to a cytokine or other ligand that provides targeting specificity, and/or to a protein transducing domain (see

generally U.S. applications Ser. No. 60/111,701 and 60/056,713, both of which are herein incorporated by reference, for examples of such transducing sequences) to facilitate passage across cellular membranes and can 5 optionally be engineered to include nuclear localization sequences.

### 5.3 ANTIBODIES TO NHP PRODUCTS

Antibodies that specifically recognize one or more 10 epitopes of a NHP, or epitopes of conserved variants of a NHP, or peptide fragments of a NHP are also encompassed by the invention. Such antibodies include but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab 15 fragments,  $F(ab')_2$  fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above.

The antibodies of the invention may be used, for example, in the detection of NHP in a biological sample and may, 20 therefore, be utilized as part of a diagnostic or prognostic technique whereby patients may be tested for abnormal amounts of NHP. Such antibodies may also be utilized in conjunction with, for example, compound screening schemes for the evaluation of the effect of test compounds on expression 25 and/or activity of a NHP gene product. Additionally, such antibodies can be used in conjunction gene therapy to, for example, evaluate the normal and/or engineered NHP-expressing cells prior to their introduction into the patient. Such antibodies may additionally be used as a method for the 30 inhibition of abnormal NHP activity. Thus, such antibodies may, therefore, be utilized as part of treatment methods.

For the production of antibodies, various host animals may be immunized by injection with the NHP, an NHP peptide (e.g., one corresponding the a functional domain of an NHP), 35 truncated NHP polypeptides (NHP in which one or more domains have been deleted), functional equivalents of the NHP or mutated variant of the NHP. Such host animals may include but

are not limited to pigs, rabbits, mice, goats, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's adjuvant (complete and 5 incomplete), mineral salts such as aluminum hydroxide or aluminum phosphate, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*.

10 Alternatively, the immune response could be enhanced by combination and or coupling with molecules such as keyhole limpet hemocyanin, tetanus toxoid, diphtheria toxoid, ovalbumin, cholera toxin or fragments thereof. Polyclonal antibodies are heterogeneous populations of antibody molecules 15 derived from the sera of the immunized animals.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, can be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, 20 but are not limited to, the hybridoma technique of Kohler and Milstein, (1975, *Nature* 256:495-497; and U.S. Patent No. 4,376,110), the human B-cell hybridoma technique (Kosbor et al., 1983, *Immunology Today* 4:72; Cole et al., 1983, *Proc. Natl. Acad. Sci. USA* 80:2026-2030), and the EBV-hybridoma 25 technique (Cole et al., 1985, *Monoclonal Antibodies And Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated *in vitro* or *in vivo*. 30 Production of high titers of mAbs *in vivo* makes this the presently preferred method of production.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison et al., 1984, *Proc. Natl. Acad. Sci.*, 81:6851-6855; Neuberger et al., 1984, *Nature*, 35 312:604-608; Takeda et al., 1985, *Nature*, 314:452-454) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a

human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a 5 human immunoglobulin constant region. Such technologies are described in U.S. Patents Nos. 6,075,181 and 5,877,397 and their respective disclosures which are herein incorporated by reference in their entirety.

Alternatively, techniques described for the production of 10 single chain antibodies (U.S. Patent 4,946,778; Bird, 1988, Science 242:423-426; Huston et al., 1988, Proc. Natl. Acad. Sci. USA 85:5879-5883; and Ward et al., 1989, Nature 334:544-546) can be adapted to produce single chain antibodies against 15 NHP sequence products. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, such fragments 20 include, but are not limited to: the  $F(ab')_2$  fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the  $F(ab')_2$  fragments. Alternatively, Fab expression libraries may be constructed (Huse et al., 1989, 25 Science, 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Antibodies to a NHP can, in turn, be utilized to generate anti-idiotype antibodies that "mimic" a given NHP, using 30 techniques well known to those skilled in the art. (See, e.g., Greenspan & Bona, 1993, FASEB J 7(5):437-444; and Nissinoff, 1991, J. Immunol. 147(8):2429-2438). For example antibodies which bind to a NHP domain and competitively inhibit the binding of NHP to its cognate receptor can be used 35 to generate anti-idiotypes that "mimic" the NHP and, therefore, bind and activate or neutralize a receptor. Such anti-idiotypic antibodies or Fab fragments of such anti-

idiotypes can be used in therapeutic regimens involving a NHP signaling pathway.

The present invention is not to be limited in scope by the specific embodiments described herein, which are intended 5 as single illustrations of individual aspects of the invention, and functionally equivalent methods and components are within the scope of the invention. Indeed, various modifications of the invention, in addition to those shown and described herein will become apparent to those skilled in the 10 art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims. All cited publications, patents, and patent applications are herein incorporated by reference in their entirety.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising at least 24 contiguous bases of nucleotide sequence first 5 disclosed in the NHP sequence described in SEQ ID NO: 3.
2. An isolated nucleic acid molecule comprising a nucleotide sequence that:
  - (a) encodes the amino acid sequence shown in SEQ ID 10 NO: 4; and
  - (b) hybridizes under stringent conditions to the nucleotide sequence of SEQ ID NO: 3 or the complement thereof.
- 15 3. An isolated nucleic acid molecule encoding the amino acid sequence described in SEQ ID NO: 6.
4. An isolated nucleic acid molecule encoding the amino acid sequence described in SEQ ID NO: 2.

20

## SEQUENCE LISTING

<110> Walke, D. Wade  
 Wilganowski, Nathaniel L.  
 Donoho, Gregory  
 Turner, C. Alexander Jr.

<120> Novel Human Proteases and Polynucleotides Encoding the Same  
 <130> LEX-0114-PCT  
 <150> US 60/174,686  
 <151> 2000-01-06  
 <160> 7  
 <170> FastSEQ for Windows Version 4.0

<210> 1  
 <211> 654  
 <212> DNA  
 <213> Homo sapien

<400> 1  
 atgtggacag ctgtgattgg aactaataat atacatggac gctatcctca taccoagaag 60  
 ataaaaatata aagcaatcat tattcatcca aacttcattt tggaatctta tgtaaatgat 120  
 attgcactttt ttcacttaaa aaaagcagtg aggtataatg actatattca gcctattgc 180  
 ctaccttttgc atgttttcca aatcctggac ggaaacacaa agtgttttat aagtggctgg 240  
 ggaagaacaa aagaagaagg taacgctaca aatattttac aagatgcaga agtgcattat 300  
 atttctcgag agatgtgtaa ttctgagagg agttatgggg gaataattcc taacacttca 360  
 ttttgcag gtgatgaaga tggagctttt gatacttgca ggggtgacag tgggggacca 420  
 ttaatgtgct acttaccaga atataaaaga tttttgtaa tgggattac cagttacgga 480  
 catggctgtg gtcgaagagg ttttcctggt gtctatattt ggccatcctt ctaccaaaag 540  
 tggctgacag agcatttctt ccatgcaagc actcaaggca tacttactat aaatatttta 600  
 cgtggccaga tcctcatagc tttatgtttt gtcatcttac tagcaacaac ataa 654

<210> 2  
 <211> 217  
 <212> PRT  
 <213> Homo sapien

<400> 2  
 Met Trp Thr Ala Val Ile Gly Thr Asn Asn Ile His Gly Arg Tyr Pro  
 1 5 10 15  
 His Thr Lys Lys Ile Lys Ile Lys Ala Ile Ile Ile His Pro Asn Phe  
 20 25 30  
 Ile Leu Glu Ser Tyr Val Asn Asp Ile Ala Leu Phe His Leu Lys Lys  
 35 40 45  
 Ala Val Arg Tyr Asn Asp Tyr Ile Gln Pro Ile Cys Leu Pro Phe Asp  
 50 55 60  
 Val Phe Gln Ile Leu Asp Gly Asn Thr Lys Cys Phe Ile Ser Gly Trp  
 65 70 75 80  
 Gly Arg Thr Lys Glu Glu Gly Asn Ala Thr Asn Ile Leu Gln Asp Ala  
 85 90 95  
 Glu Val His Tyr Ile Ser Arg Glu Met Cys Asn Ser Glu Arg Ser Tyr  
 100 105 110  
 Gly Gly Ile Ile Pro Asn Thr Ser Phe Cys Ala Gly Asp Glu Asp Gly  
 115 120 125  
 Ala Phe Asp Thr Cys Arg Gly Asp Ser Gly Gly Pro Leu Met Cys Tyr  
 130 135 140  
 Leu Pro Glu Tyr Lys Arg Phe Phe Val Met Gly Ile Thr Ser Tyr Gly  
 145 150 155 160

His Gly Cys Gly Arg Arg Gly Phe Pro Gly Val Tyr Ile Gly Pro Ser  
 165 170 175  
 Phe Tyr Gln Lys Trp Leu Thr Glu His Phe Phe His Ala Ser Thr Gln  
 180 185 190  
 Gly Ile Leu Thr Ile Asn Ile Leu Arg Gly Gln Ile Leu Ile Ala Leu  
 195 200 205  
 Cys Phe Val Ile Leu Leu Ala Thr Thr  
 210 215

&lt;210&gt; 3

&lt;211&gt; 1047

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 3

atgcggctgg	ggctcctgag	cgtggcgttg	ttgtttgtgg	ggagctctca	cttayactca	60
gaccactact	cgcctctgg	aaggcacagg	ctcgccccc	cgccggaaacc	ggcggttagt	120
tcccagcagg	ctgaggccgt	ccgcaagagg	ctccggcgcc	ggagggagggg	agggggcgcat	180
gcaaaggatt	gtggAACAGC	accgcttaag	gatgtgttgc	aagggtotcg	gattataggg	240
ggcaccgaag	cacaagctgg	cgcattggccg	tgggtggtga	gcctgcagat	taaatatggc	300
cgtgttcttg	ttcatgtatg	tggggaaacc	ctagtgagag	agaggtgggt	cctcacagct	360
gcccactgca	ctaaagacrc	tagcgatct	ttaatgttga	cagctgtat	tggactaat	420
aatatacatg	gacgctatcc	tcataccaag	aagataaaaa	ttaaagcaat	cattattcat	480
ccaaacttca	tttggaaatc	ttatgttaat	gatattgcac	tttttcaatt	aaaaaaagca	540
gtgaggata	atgactat	tcagcctatt	tgcctacctt	ttgatgtttt	ccaaatcctg	600
gacggaaaca	caaagtgttt	tataagtggc	tggggaaagaa	caaaaagaaga	aggtaacgct	660
acaaatattt	tacaagatgc	agaagtgcac	tatatttctc	gagagatgtg	taattctgag	720
aggagttatg	ggggaaataat	tcctaacact	tcattttgtg	caggtgatga	agatggagct	780
tttgatactt	gcaggggtga	cagtggggga	ccattaatgt	gctacttacc	agaatataaa	840
agatttttg	taatgggaat	taccagttac	ggacatggct	gtggcgaag	aggtttcct	900
ggtgtctata	ttgggccatc	cttctaccaa	aagtggatga	cagacattt	cttccatgca	960
agcactcaag	gcatacttac	tataaatatt	ttacgtggcc	agatcctcat	agctttatgt	1020
tttgtcatct	tactagcaac	aacataa				1047

&lt;210&gt; 4

&lt;211&gt; 348

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 4

Met Arg Leu Gly Leu Leu Ser Val Ala Leu Leu Phe Val Gly Ser Ser						
1	5	10	15			
His Leu Tyr Ser Asp His Tyr Ser Pro Ser Gly Arg His Arg Leu Gly						
20	25	30				
Pro Ser Pro Glu Pro Ala Ala Ser Ser Gln Gln Ala Glu Ala Val Arg						
35	40	45				
Lys Arg Leu Arg Arg Arg Glu Gly Gly Ala His Ala Lys Asp Cys						
50	55	60				
Gly Thr Ala Pro Leu Lys Asp Val Leu Gln Gly Ser Arg Ile Ile Gly						
65	70	75	80			
Gly Thr Glu Ala Gln Ala Gly Ala Trp Pro Trp Val Val Ser Leu Gln						
85	90	95				
Ile Lys Tyr Gly Arg Val Leu Val His Val Cys Gly Gly Thr Leu Val						
100	105	110				
Arg Glu Arg Trp Val Leu Thr Ala Ala His Cys Thr Lys Asp Ala Ser						
115	120	125				
Asp Pro Leu Met Trp Thr Ala Val Ile Gly Thr Asn Asn Ile His Gly						
130	135	140				
Arg Tyr Pro His Thr Lys Lys Ile Lys Ile Lys Ala Ile Ile Ile His						
145	150	155	160			
Pro Asn Phe Ile Leu Glu Ser Tyr Val Asn Asp Ile Ala Leu Phe His						
165	170	175				
Leu Lys Lys Ala Val Arg Tyr Asn Asp Tyr Ile Gln Pro Ile Cys Leu						

180	185	190
Pro Phe Asp Val Phe Gln Ile Leu Asp Gly Asn Thr Lys	Cys Phe Ile	
195 200	205	
Ser Gly Trp Gly Arg Thr Lys Glu Glu Gly Asn Ala Thr Asn	Ile Leu	
210 215	220	
Gln Asp Ala Glu Val His Tyr Ile Ser Arg Glu Met Cys Asn Ser	Glu	
225 230	235	240
Arg Ser Tyr Gly Gly Ile Ile Pro Asn Thr Ser Phe Cys Ala	Gly Asp	
245 250	255	
Glu Asp Gly Ala Phe Asp Thr Cys Arg Gly Asp Ser Gly	Gly Pro Leu	
260 265	270	
Met Cys Tyr Leu Pro Glu Tyr Lys Arg Phe Phe Val Met Gly Ile	Thr	
275 280	285	
Ser Tyr Gly His Gly Cys Gly Arg Arg Gly Phe Pro Gly Val	Tyr Ile	
290 295	300	
Gly Pro Ser Phe Tyr Gln Lys Trp Leu Thr Glu His Phe Phe His	Ala	
305 310	315	320
Ser Thr Gln Gly Ile Leu Thr Ile Asn Ile Leu Arg Gly Gln Ile	Leu	
325 330	335	
Ile Ala Leu Cys Phe Val Ile Leu Leu Ala Thr Thr		
340 345		

&lt;210&gt; 5

&lt;211&gt; 867

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 5

atgcaaaatt gtggaacacgc accgcttaag gatgtgttgc aagggtctcg gattataggg	60
ggcacccgaag cacaaggctgg cgcacggccg tgggtgggtga gcctgcagat taaaatatggc	120
cgtgttcttg ttcatgtatg tgggggaacc ctagtgagag agaggtgggt cctcacagct	180
gcccactgca ctaaagacrc tagcgatct ttaatgtgga cagctgtgat tggaaactaat	240
aatatacatg gacgctatcc tcataccaag aagataaaaa ttaaagcaat cattattcat	300
ccaaacttca ttttggaaatc ttatgtaaat gatattgcac ttttcactt aaaaaaagca	360
gtgagggtata atgactatat tcagcctatt tgcc tacctt ttgatgtttt ccaaattcctg	420
gacggaaaca caaagtgttt tataagtggc tggggaaagaa caaaaagaaga aggttaacgct	480
acaaatattt tacaagatgc agaagtgcat tatatttctc gagagatgtg taattctgag	540
aggagttatg ggggaataat tcctaacact tcattttgtg caggtgatga agatggagct	600
tttgataactt gcaggggtga cagtggggga ccattaatgt gctacttacc agaatataaa	660
agatttttg taatgggaat taccagttac ggacatggct gtggcgaag aggtttccct	720
ggtgtctata ttgggccatc cttctaccaa aagtggctga cagagcattt cttccatgca	780
agcaactcaag gcatacttac tataaatatt ttacgtggcc agatcctcat agctttatgt	840
tttgcatct tactagcaac aacataa	867

&lt;210&gt; 6

&lt;211&gt; 288

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 6

Met Gln Asn Cys Gly Thr Ala Pro Leu Lys Asp Val Leu Gln Gly Ser	
1 5 10 15	
Arg Ile Ile Gly Gly Thr Glu Ala Gln Ala Gly Ala Trp Pro Trp Val	
20 25 30	
Val Ser Leu Gln Ile Lys Tyr Gly Arg Val Leu Val His Val Cys Gly	
35 40 45	
Gly Thr Leu Val Arg Glu Arg Trp Val Leu Thr Ala Ala His Cys Thr	
50 55 60	
Lys Asp Ala Ser Asp Pro Leu Met Trp Thr Ala Val Ile Gly Thr Asn	
65 70 75 80	
Asn Ile His Gly Arg Tyr Pro His Thr Lys Lys Ile Lys Ile Lys Ala	
85 90 95	
Ile Ile His Pro Asn Phe Ile Leu Glu Ser Tyr Val Asn Asp Ile	

100	105	110
Ala Leu Phe His Leu Lys Lys	Ala Val Arg Tyr Asn Asp	Tyr Ile Gln
115	120	125
Pro Ile Cys Leu Pro Phe Asp Val Phe Gln	Ile Leu Asp Gly Asn Thr	
130	135	140
Lys Cys Phe Ile Ser Gly Trp Gly Arg Thr	Lys Glu Glu Gly Asn Ala	
145	150	155
Thr Asn Ile Leu Gln Asp Ala Glu Val His	Tyr Ile Ser Arg Glu Met	
165	170	175
Cys Asn Ser Glu Arg Ser Tyr Gly Gly	Ile Ile Pro Asn Thr Ser Phe	
180	185	190
Cys Ala Gly Asp Glu Asp Gly Ala Phe Asp	Thr Cys Arg Gly Asp Ser	
195	200	205
Gly Gly Pro Leu Met Cys Tyr Leu Pro Glu	Tyr Lys Arg Phe Phe Val	
210	215	220
Met Gly Ile Thr Ser Tyr Gly His	Gly Cys Gly Arg Arg Gly Phe Pro	
225	230	235
Gly Val Tyr Ile Gly Pro Ser Phe Tyr	Gln Lys Trp Leu Thr Glu His	
245	250	255
Phe Phe His Ala Ser Thr Gln Gly	Ile Leu Thr Ile Asn Ile Leu Arg	
260	265	270
Gly Gln Ile Leu Ile Ala Leu Cys Phe Val	Ile Leu Ala Thr Thr	
275	280	285

&lt;210&gt; 7

&lt;211&gt; 1286

&lt;212&gt; DNA

&lt;213&gt; Homo sapien

&lt;400&gt; 7

ttcttccatt tcaggtgtcg	tgaaaagctt gaattcggcg	cgccagatat cacacgtgcc	60
aaggggctgg ctcgcccaca	tcttgcac cagcctccaa	aatgcggctg gggctcctga	120
gcgtggcgct gttgtttgtg	gggagctctc	acttatactc agaccactac	180
gaaggcacag gtcggcccc	tcgcccgaac	cggcggctag ttcccagcag	240
tccgcaagag gtcggcgcc	cggagggagg	gaggggcgcga tgcaaaggat	300
caccgcttaa ggatgtgtg	caagggtctc	ggattatagg gggcacccgaa	360
gcgcatggcc gtgggtggtg	agcctgcaga	gcacaagctg tttcatgtat	420
gtggggaaac cctagtgaga	gagaggtggg	tcctcacagc tgccactgc actaaagacg	480
ctagcgatcc tttatgtgg	acagctgtg	ttgaaactaa taatatacat	540
ctcataccaa gaagataaaa	attaaagcaa	tttttggat	600
cttatgtaaa tgatattgca	cttttact	taaaaaaagc agtggat	660
ttcagccat ttgcctacct	tttgatgttt	tccaaatcct ggacggaaac	720
ttataagtgg ctggggaga	acaaaagaag	acaaagtgtt	780
cagaagtgca ttatatttct	cgagagatgt	ttttgataact	840
ttcctaacac ttcatttgt	gcaggtgatg	ttttgttgc	900
acagtgggg accattaatg	tgctacttac	ttttgttgc	960
ttaccaggta cggacatggc	tgtggtcgaa	ttttgttgc	1020
ccttctacca aaagtggctg	gaggtttcc	ttttgttgc	1080
ctataaatat ttacgtggc	tggtgtctat	ttttgttgc	1140
caacataaag aaattctgaa	ttttgttgc	ttttgttgc	1200
ttctatataa tgaacatcat	ttattctct	ttttgttgc	1260
tgtgaacatt ttatggcata	agcaattaat	ttttgttgc	1286

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 01/00548

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 C12N15/57

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

BIOSIS, EMBASE, CHEM ABS Data, STRAND, EPO-Internal

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category <sup>o</sup>	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE STRAND 'Online! AC008121, 26 July 1999 (1999-07-26) MUZNY, D.M. ET AL.: "Homo sapiens chromosome 12 clone RP11-407N8" XP002166333 *100% identity in 275 bp overlap (total 1047 bp)* abstract</p> <p>-----</p> <p>DATABASE STRAND 'Online! AC013244, 9 November 1999 (1999-11-09) MUZNEY, D.M. ET AL. : "Homo sapiens chromosome 12 clone RP11-60E8" XP002166334 *100% identity in 275 bp overlap (total 1047 bp). abstract</p> <p>-----</p> <p>-----</p>	1
X		1

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

<sup>o</sup> Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the International filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

2 May 2001

Date of mailing of the international search report

17/05/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Hiltenbrand, G

## INTERNATIONAL SEARCH REPORT

Int'l Application No  
PCT/US 01/00548

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE STRAND 'Online! AC009727, 1 September 1999 (1999-09-01) MUZNEY, D.M. ET AL. : "Homo sapiens chromosome 12 clone RP11-533J15" XP002166335 *100% identity in 275 bp overlap (total 1047 bp)* abstract ---	1
A	WO 99 57274 A (INCYTE PHARMA INC ;PATTERSON CHANDRA (US); CORLEY NEIL C (US); GUE) 11 November 1999 (1999-11-11) the whole document ---	1
A	WO 99 53078 A (GENENCOR INT) 21 October 1999 (1999-10-21) the whole document ---	1
A	WO 99 36550 A (INCYTE PHARMA INC ;CORLEY NEIL C (US); YUE HENRY (US); BANDMAN OLG) 22 July 1999 (1999-07-22) the whole document ---	1

## INTERNATIONAL SEARCH REPORT

Information on patent family members

Int'l Application No
PCT/US 01/00548

Patent document cited in search report	Publication date	Patent family member(s)			Publication date
WO 9957274	A 11-11-1999	US 6171790 B	AU 3768199 A	EP 1076702 A	09-01-2001 23-11-1999 21-02-2001
WO 9953078	A 21-10-1999	AU 3643299 A	BR 9909639 A	EP 1071792 A	01-11-1999 19-12-2000 31-01-2001
WO 9936550	A 22-07-1999	US 6203979 B	AU 2113599 A	EP 1045913 A	20-03-2001 02-08-1999 25-10-2000